

Abstract Title Page
Not included in page count.

**Title: Analysis of Textual Features of a New Reading Comprehension Assessment:
MOCCA**

Authors and Affiliations:

Ben Seipel, California State University, Chico
Gina Biancarosa, University of Oregon, Center on Teaching & Learning
Sarah Carlson, University of Oregon, Center on Teaching & Learning
Mark Davison, University of Minnesota, Twin Cities

Abstract Body

Background / Context:

Previous research has established two types of struggling readers: those who struggle with lower-level reading skills and those who struggle with higher-level reading skills (Cain & Oakhill, 2006; Perfetti, 2007). The latter group is commonly termed *poor comprehenders*: readers who exhibit poor comprehension compared to peers with similar word-reading skills and vocabulary (e.g., Cain & Oakhill, 1999, 2006; Carlson, Seipel, & McMaster, 2011, 2014; Rapp et al., 2007). Moreover, research has revealed that poor comprehenders exhibit difficulty with causally coherent inferences (e.g., McMaster et al., 2012; Trabasso & van den Broek, 1985).

Causally coherent inferences require synthesis of why an event occurs based on relevant goals and subgoals previously identified in the text *and* generate missing information from background knowledge consistent with this synthesis. Although poor comprehenders *do* make causally coherent inferences, they do not make them as *consistently* as good comprehenders.

Instead, when poor comprehenders do not make such inferences, they often use other types of comprehension processes that are strategic and useful, but that fail to fill the causal gap in the text. Specifically, poor comprehenders tend either to paraphrase (i.e., rephrase prior information from the text but do not generate missing information) or to make lateral connections (i.e., make an elaborative inference or a personal association, which use background knowledge but are *not causally coherent* with the text). These trends have been found repeatedly with intermediate grade readers (i.e., Grades 3-5; e.g., McMaster et al., 2012; Rapp et al., 2007). Given the consistency of such findings, if we were able to use an assessment to distinguish which processes poor comprehenders are relying on, we would be able deliver more targeted instruction.

Nonetheless, measures of reading comprehension remain incapable of distinguishing among the processes poor comprehenders use. Rather, existing reading comprehension measures focus almost exclusively on comprehension as *product* and provide little to no information about *how* a reader arrived at that product: the comprehension *process*. As a result, there have been repeated calls for measures reflecting individual differences in comprehension processes (Klingner, 2004; Pearson & Hamm, 2005; RAND Reading Study Group, 2002). Furthermore, the research on reading comprehension processes differences has relied almost exclusively on think alouds, which require readers to verbally report what they think as they read. However, think alouds are too laborious and time-consuming to be practical assessments for schools to use.

As a result, more practical measures of the reading comprehension process, to date, have been extremely specific and either target specific populations, such as adult readers (e.g., Hannon & Daneman, 2001), look at inferences in the presence or absence of supportive illustrations (Pike, Barnes, & Barron, 2010), or use texts that are a series of logical, relational statements rather than more common narrative and expository forms (e.g., August, Francis, Hsu, & Snow, 2006). Critically, none offer diagnostic information about what poor comprehenders *are doing* when they read, just what they are *not doing successfully*.

Development of the MOCCA

The original Multiple-choice Online Cloze Comprehension Assessment (MOCCA) was designed to differentiate among poor comprehenders in terms of the types of comprehension processes they use during reading (cf. Carlson et al., 2014). MOCCA consists of 40 items, which use short narrative texts (seven sentences long) written around a fourth grade Flesch-Kincaid Grade Level (Kincaid, Fishburne, Rogers, & Chissom, 1975). Each item (i.e., text) is a discourse-level maze task in which the sixth sentence is deleted. Readers read each text and choose among four multiple-choice response types to complete the missing sentence (see Figure 1 in Appendix B for example). The four multiple-choice responses are: (1) a *causally coherent inference*, (2) a *paraphrase*, (3) a *lateral connection*, and (4) a *local bridging inference*. Causally coherent inferences are always the best response to complete the story in a causally coherent manner. Paraphrases are always incorrect, but mimic what one group of poor comprehenders

tend to do when they do not make causally coherent inferences (i.e., paraphrase) during reading. Lateral connections are also always incorrect, but mimic what another group of poor comprehenders tend to do when they do not make causally coherent inferences (i.e., make an elaborative inference or make a personal association, neither of which completes the causal chain of events). Note that readers may draw on background knowledge to make either causally coherent inferences or lateral connections. What distinguishes the two is whether they close the causal chain of events or not. Finally, local bridging inferences are also incorrect, and while they are a type of inference readers *can* make and are easier to make than other types of inferences, they do *not* tend to distinguish between types of comprehenders.

Each of the original MOCCA narrative texts were written so the causal structure of each text (i.e., plot, nature of events) was developed around a main goal that motivates subgoals and events in the text (e.g., Trabasso, van den Broek, & Suh, 1989). Each text begins with a title. The first sentence is an introductory sentence and introduces the main character of the story. The second sentence generally introduces the main goal of the story or may further plot development. In the third, fourth, and fifth sentences, a subgoal of the story is introduced and detailed. Here the subgoal may introduce another challenge (e.g., character needs to go to the store buy butter to make the cookies) or simply detail additional steps to completing the main goal (e.g., character needs to grab a pencil to do homework). The sixth sentence is always the deleted sentence and resolves the goal or is required to resolve the goal of the story. The last (i.e., seventh) sentence of each text concludes the story. Answer choices were developed based on different types of comprehension responses identified in previous think-aloud research (McMaster et al., 2012; Rapp et al., 2007). In addition to the macrostructure of the test items, other textual features are of interest.

Previous research has indicated that the ability of a reader to understand any text is based on the reader's ability and experience (i.e., phonemic awareness, decoding, lexical access, vocabulary knowledge, fluency, inference ability, and working memory), a reader's goals for reading a given text, and the text itself (i.e., readability, genre, titles) (Kozminsky, 1977; van den Broek, Lorch, Linderholm, & Gustafson, 2001). Such textual features are important to note in traditional reading comprehension tests where students generally read a limited number of passages and then answer multiple literal and inferential questions about the text passage. The textual features of MOCCA, however, are of dual interest because each test item is designed to diagnose issues with comprehension and each item is one complete reading passage. Therefore, it is important to understand how various textual features of texts may affect comprehension.

Several textual features have been well studied: readability, titles, goal structures, goal location, explicitness of the goal, cohesion of text through the use of connectives, secondary agents, and even gender stereotypes based characters' names (Garnham, Oakhill, & Reynolds, 2002). Readability is a metric of text difficulty usually based on the number of syllables, words, and sentences in a text. Common formulae for readability include Flesch-Kincaid Grade Level, Reading Ease (Kincaid, Fishburne, Rogers, & Chissom, 1975), and Dale-Chall Readability (Chall, 1948). Although readability can be useful in determining whether a text is too difficult for a particular age/grade level, it is not a good predictor of whether or not students actually understand text (Wait, 1987 as cited by Zamanian & Heydari, 2012). Titles affect comprehension as well. They act as advanced organizers (Ausubel, 1968; Haviland & Clark, 1974) and can bias comprehension (Kozminsky, 1977). Textual features such as physical layout, text structure, and use of connectives also influence comprehension (e.g., Graesser, McNamara, Louwerse, & Cai, 2004). In MOCCA, each text/item has a narrative story that is structured the same. However, the place of the main goal statement, main idea, and its explicitness of the main goal vary by item. These features also can influence comprehension of a story (Beishuizen, Asscher, Prinsen, &

Elshout-Mohr, 2003; Wang, 2009). Given the originality of this type of reading comprehension assessment in structure, its diagnostic nature, and its forthcoming revisions, it is imperative to determine which textual features affect item-level statistics.

Purpose / Objective / Research Question / Focus of Study:

The purpose of the current study was to determine which textual features are correlated with and affect item difficulty (proportion correct, item discrimination, and point-biserial correlations for the original MOCCA test items. This information will be used to develop item-writing guidelines, edit original test items, and write new test items.

Setting:

The original MOCCA test data were collected in 3rd, 4th, and 5th grade classrooms during group administration. Classrooms from suburban districts of a larger upper-Midwestern city agreed to participate in the original data collection.

Population / Participants / Subjects:

The population of interest is all 40 original MOCCA test items. The average grade-level readability (Flesch-Kincaid grade level) of the test items was 4.43 (SD=.86; min. 2.30, max. 6.40 grade levels). The student sample from which test item difficulty, proportion correct, and item discrimination was determined based on 192 3rd, 4th, and 5th grade students.

Intervention / Program / Practice:

The results of this study will be used to inform the writing of new MOCCA test items to better differentiate between good and poor comprehenders, and the better differentiate between the two types of struggling comprehenders.

Research Design:

Original MOCCA test items were coded for various textual features that have been previously identified as affecting item difficulty (e.g., readability) as well as for textual features endemic to MOCCA test items (e.g., location of main goal statement within a MOCCA item). Disagreements about the presence of a textual feature in any one MOCCA item (i.e., explicit main goal, emotional resolution) were resolved by discussion. Textual features included for coding include: Flesch-Kincaid grade level; whether or not the final sentence was compound or simple in grammatical structure (i.e., used a conjunction to combine two or more clauses; whether or not the final sentence was complex in grammatical structure (i.e., included embedded clause), in which sentence(s) the main goal was indicated; whether or not the main goal was introduced in one sentence or inferred over two or more sentences; whether or not the main goal was an avoidance goal (e.g., character wanted to avoid doing a chore); from which sentence the paraphrase response was modeled; whether or not the paraphrase response was a combination of two or more sentences; whether or not the test item included a secondary agent or not (note that a primary agent could be two characters who acted in unison to complete a single goal); whether or not the emotional valence in the final sentence was explicit or not; whether or not the emotion in the final was inferred; whether the emotion in the final sentence was positive or negative in valence; whether or not the emotion in the final sentence was inferred based on a verb (e.g., cried, smiled); whether or not the subgoal presented an additional challenge to meeting the main goal or not; whether or not the title of the item mentions a character or not; whether or not

the title indicated the situation or not; whether the item has an animal character or not; and whether the title indicated that there was animal in the story or not.

Item difficulty (proportion of students answering the item correctly with the causally coherent response), item discrimination, and point-biserial of original MOCCA items were previously calculated and published (Carlson et al., 2014). We calculated correlations between textual features and the original item difficulty statistics while controlling for item position, as item position differentially affects item difficulty (Debeer & Janssen, 2013).

Data Collection and Analysis:

Student MOCCA data were originally collected during group test settings as described above. Textual features of the original MOCCA items were determined during a two-day item-revising workshop.

Findings / Results:

Correlations between original MOCCA item features and item statistics are displayed in Table 1 (see Appendix B). Most textual features of the original MOCCA items were uncorrelated with item difficulty (proportion of items correct), item discrimination, and point-biserial. A few item features were correlated with proportion correct. Specifically, the presence of a secondary agent (with his/her/their own goals) made an item more difficult to answer correctly. In addition, the explicit mention of an animal in the item/story title made an item easier to answer correctly. A final sentence with a verb that indicated emotion (e.g., he cried, she smiled) also marginally made an item easier to answer correctly. A final sentence with a positive (i.e., happy) ending made an item marginally more difficult to answer correctly. A few item features were also correlated with item discrimination; the presence of a compound final sentence, the presence of a secondary agent, and explicit emotion in the final sentence was better for discriminating between good and poor comprehenders. A final sentence with a verb that indicated emotion made an item less likely to differentiate between good and poor comprehenders. Finally, no item features were correlated with point-biserial. Some item features (e.g., neutral ending, gender neutral names, unnamed characters) only existed for one or two items and made further analysis untenable.

Conclusions:

The purpose of this study was to determine which textual features of the original MOCCA test items contribute to test item difficulty (proportion of students answering the item correctly with the causally coherent response), item discrimination, and point-biserial. This information will be used to develop item-writing guidelines, edit original test items, and write new test items. Findings from these analyses will specifically allow us to develop new test items at various grade levels (3rd-5th grades). That is, we plan to systematically vary the presence of animals in item/stories and in their titles, balance items/stories regarding the final sentence's emotion (i.e., positive/ negative, inferred) and explicitness, and vary the presence of a secondary agent with different or conflicting goals of the primary agent. There are limitations to the current study. First, the number of test takers at each grade level is relatively low for the desired item analysis. Second, as indicated, successful comprehension of text requires integration of text features, reader ability, and reader goal. This study only examined the textual features that correlated with comprehension. Future item development, test administration, and IRT analysis will aid in determining interactive effects of reader ability, reader goal, and item features.

Appendices

Not included in page count.

Appendix A. References

- August, D., Francis, D. J., Hsu, H.-Y. A., & Snow, C. E. (2006). Assessing reading comprehension in bilinguals. *Elementary School Journal*, 107, 221–239.
- Beishuizen, J., Asscher, J., Prinsen, F., & Elshout-Mohr, M. (2003). Presence and place of main ideas and examples in study texts. *British Journal of Educational Psychology*, 73, 291–316.
- Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing*, 11, 489–503. doi: 10.1023/a:1008084120205
- Cain, K., & Oakhill, J. V. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology*, 76, 683–696. doi: 10.1348/000709905x67610
- Carlson, S. E., Seipel, B., & McMaster, K. (2014). Development of a new reading comprehension assessment: Identifying different types of comprehenders. *Learning and Individual Differences*, 32, 40–53. DOI: 10.1016/j.lindif.2014.03.003
- Dale, E., & Chall, J. (1948). A formula for predicting readability". *Educational Research Bulletin* 27, 11–20+28.
- Debeer, D. & Janssen, R. (2013). Modeling item-position effects within an IRT framework. *Journal of Education Measurement*, 50(2), 164–185.
- Garnham, A., Oakhill, J. & Reynolds D. (2002). Are inferences from stereotyped role names to characters' gender made elaboratively? *Memory & Cognition*, 30(3), 439–446.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371–395. doi: 10.1037/0033-295x.101.3.371
- Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology*, 93, 103–128. doi: 10.1037/0022-0663.93.1.103
- Kincaid, J. P., Fishburne, L. R. P., Jr., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for Navy enlisted personnel*. (Research Branch Report No. 8-75). Millington, TN: Naval Air Station Memphis (75). Retrieved from <http://library.ucf.edu/Web/purl.asp?dpid=DP0008946>
- Klingner, J. K. (2004). The science of professional development. *Journal of Learning Disabilities*, 37, 248–255. doi: 10.1177/00222194040370031001
- Kozminksy, E. (1977). Altering comprehension: The effect of biasing titles on text comprehension. *Memory & Cognition*, 5(4), 482–490.
- Leslie, L., & Caldwell, J. (2001). *Qualitative reading inventory- 3*. New York: Longman.
- McMaster, K. L., den Broek, P. v., A. Espin, C., White, M. J., Rapp, D. N., Kendeou, P., . . . Carlson, S. (2012). Making the right connections: Differential effects of reading intervention for subgroups of comprehenders. *Learning and Individual Differences*, 22, 100–111. doi: 10.1016/j.lindif.2011.11.017
- Pearson, P. D., & Hamm, D. N. (2005). The assessment of reading comprehension: A review of practices-past, present, and future. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 13–69). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11, 357–383. doi: 10.1080/10888430701530730

- Pike, M. M., Barnes, M. A., & Barron, R. W. (2010). The role of illustrations in children's inferential comprehension. *Journal of Experimental Child Psychology*, 105, 243–255. doi: 10.1016/j.jecp.2009.10.006
- RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND. Retrieved from <http://uolibraries.worldcat.org/oclc/50851156>
- Rapp, D. N., Broek, P. v. d., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading*, 11, 289–312. doi: 10.1080/10888430701530417
- Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, 24, 612–630. doi: 10.1016/0749-596x(85)90049-x
- Trabasso, T., van den Broek, P., & Suh, S. Y. (1989). Logical necessity and transitivity of causal relations in stories. *Discourse Processes*, 12, 1–25. doi: 10.1080/01638538909544717
- van den Broek, P., Lorch, R., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, 29, 1081–1087. doi: 10.3758/bf03206376
- Wang, D. (2009). Factors affecting the comprehension of global and local main idea. *Journal of College Reading and Learning*, 39 (2), 34-52.
- Zamanian, M. & Heydari, P. (2012). Readability of texts: State of the art. *Theory and Practice in Language Studies*, 2(1). 43-53. doi:10.4304/tpls.2.1.43-53

Appendix B. Tables and Figures

Figure 1. Example MOCCA test item; examinees were instructed to determine which response best completes the story at the point “MISSING SENTENCE.”

James and the Jar

While playing in Sam’s backyard, James noticed a glass jar with a piece of paper. He wanted to see what it was, so he picked it up and unscrewed the lid. Just then, Sam ran up behind James and took the piece of paper. Sam ran to the big tree house and climbed up to get away from James. The big tree house was high up in an oak tree and had a rope ladder.

MISSING SENTENCE

James saw that the note was actually a map leading to a buried treasure!

CHOICES:

- A) James found a piece of paper, but couldn’t read it because Sam took it. (Paraphrase)
- B) James climbed the ladder and grabbed the piece of paper from Sam. (Causally Coherent)
- C) The tree house also had several big windows and a television set. (Lateral Connection)
- D) Sam pulled up the rope ladder so James couldn’t get to the paper. (Local Bridging)

Table 1. Correlations between original MOCCA item features and item statistics.

Item feature	Item statistics		
	Proportion correct	Discrimination	Point-biserial
Flesch-Kincaid grade level	-.126	-.068	-.170
Compound final sentence	-.184	.364*	.231
Complex final sentence	-.205	-.204	-.301
Compound or complex final sentence	-.260	.170	-.057
Main goal in sentence 1	.075	.200	.209
Main goal in sentence 2	.065	-.254	-.182
Main goal is combo of sentences	-.104	.078	.043
Main goal is explicit	.084	.230	.269
Subgoal introduces added challenge	-.075	.095	.040
Challenge missing in goal	-.127	-.240	-.285
Secondary agent	-.353*	.354*	.146
Emotion in final sentence is explicit	.114	.356*	.293
Emotion in final sentence is positive	-.292(*)	-.055	-.159
Emotion in final inferred	.053	-.052	.041
Emotion in final sentence is verb-based	.274(*)	-.320*	-.176
Paraphrase response sentence	.014	.234	.182
Paraphrase response sentence combo	-.030	.120	-.001
Title has character	-.220	-.088	-.165
Title has situation	.252	-.047	.039
Title has animal	.325*	.047	.225
Animal in item	.041	-.089	-.118

Note. * = $p < .05$, (*) = $p < .10$